DOCUMENT RESUME

ED 200 446 SE 034 600

AUTHOR Farmer, Walter A.; And Others
TITLE An Analysis of the Range of 2B and 3A Responses to Three Inhelder Tasks.
PUB DATE Apr 81
NOTE 36p.; Paper presented at the Annual Meeting of the National Association for Research in Science Teaching (54th, Grossinger's in the Catskills, Ellenville, NY, April 5-8, 1981).

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Cognitive Development; *Cognitive Tests; Individual Testing; *Interviews; Mathematics Education; Science Education; Secondary Education; *Secondary School Mathematics; *Secondary School Science
IDENTIFIERS Piaget (Jean); Science Education Research

ABSTRACT

Results are reported of a study designed to: (1) investigate the range of types of reasoning exhibited by ninth- and tenth-grade science and mathematics students (N=69) within 2B and 3A stages as classified by three Inhelder tasks; (2) explore the feasibility of subdividing types of responses into bands within each stage; (3) provide descriptive data on the developmental range of reasoning patterns in stages 2B and 3A; and (4) suggest tentative hypotheses related to these intellectual stages. (Author/CS)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U S DEPARTMENT OF HEALTH
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Walter A. Farmer

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

ED200446

AN ANALYSIS OF THE RANGE OF 2B AND 3A RESPONSES TO THREE INHELDER TASKS

Walter A. Farmer

Margaret A. Farrell

Stephen Rumsey

State University of New York at Albany

Annual Meeting of National Association for Research in Science Teaching, Grossinger's, Ellenville, New York, April 1981.

SE 034 600

BACKGROUND

The importance of diagnosing the level of intellectual development of secondary school students has been attested to by many researchers and practitioners (Lawson, 1979; Phillips, 1977; Shayer, 1979; Neimark, 1975, 1979) and has led to the development of diverse Piagetian-like tasks suitable for group administration. Other researchers, such as Farrell et al. (1980), Lawson (1978), and Shayer (1979), have addressed the problem of validating group paper-and-pencil tests against original Inhelder tasks (1958) and confirmed earlier findings by Renner and Stafford (1972) that a high percentage of tenth, eleventh and twelfth grade students are at the concrete operational level. Studies of college students by McKinnon and Renner (1971), Juraschek (1974) and Pluta (1980) further substantiate the existence of a large percentage of students across both age ranges and academic tracks, who do not display formal operational reasoning at the 3B level. These data on the intellectual capability of adolescents and young adults have concerned curriculum writers and instructional leaders. However, Inhelder and Piaget's (1958) own protocols remind us that making use of the dichotomy concrete/formal overlooks the diverse, developing reasoning patterns of the late concrete and early formal stages. More precise data on reasoning patterns during this transitional period would be of both

practical and theoretical significance.

OBJECTIVES

This study was designed to (1) investigate the range of types of reasoning exhibited by ninth and tenth grade science and mathematics students within 2B and 3A stages as classified by three Inhelder tasks, (2) explore the feasibility of subdividing types of responses into bands within each stage, (3) provide descriptive data on the developmental range of reasoning patterns in stages 2B and 3A, and (4) suggest tentative hypotheses related to these intellectual stages.

METHODOLOGY

In a previous study by two of the authors, 512 students enrolled in ninth and tenth grade science and mathematics classes were classified by both Sheehan's (1970) version of the Longeot subtests and a battery of three puzzles developed by Karplus, Lawson and Renner (as presented in Karplus et al., 1977). The eight participating school districts included inner-city, rural-suburban, urban-suburban and small city. The results of these group paper-and-pencil screening tests were used to select 69 subjects who then completed three Inhelder (1958) interviews--chemical combinations, flexible rods, and shadows tasks. The basis for selection was obtaining equal numbers of students across

schools and in stages 2A through 3B as classified by the group tests.

Trained interviewers employed both tape recorders and written protocol/scoring sheets during and immediately after administration of all tasks. On most occasions two interviewers were present in the room--one administering and recording and the other observing and recording. Both members of the interview team compared data and reached consensus on the classification assigned. Later a third trained interviewer checked the classifications by analyzing both tapes and written records for half of the subjects selected at random.

Next, all subjects classified as 2B or 3A were selected out by task. The two "highest" and two "lowest" 2B students on each task from each school were sorted out. The same procedure was followed with respect to 3A students. "Highest" and "lowest" were defined as reasoning patterns which matched the classical Inhelder (1958) protocols as being at the point of the next higher or next lower substage. This pool of subjects was examined to locate the two "highest" and two "lowest" 2B students and the two "highest" and two "lowest" 3A students on each task regardless of school. This final sampling produced 21 subjects whose responses provided the data base for this study. The

expected N of 24 could not be obtained because the difficulty level of the shadows task reduced that portion of the pool to three subjects classified as 3A or 3B.

Trained and experienced interviewers then analyzed the responses of the 21 subjects by listening to all relevant tapes and studying the matching scoring sheets and anecdotal records.

RESULTS

In the following pages, each task is described and the classification procedure is explained. Then the differentiating characteristics of "low" and "high" 2B and 3A subjects in the sample are delineated with specific data from one subject in each of the categories used to support and clarify the descriptive statements.

The Chemical Combinations Task. The problem posed by this task requires the subject to use a combinatorial system directly. The task is unique in that a combination of diverse factors is required if the variable results necessary for eliciting hypotheses are to be obtained. The equipment used in this study is referred to as "Experiment I" by Inhelder and Piaget (1958). It consisted of four identical-appearing bottles containing colorless, odorless liquids. Bottle one contained diluted sulphuric acid; bottle two, water; bottle

three, hydrogen peroxide solution and bottle four, sodium thiosulphate solution. A fifth bottle, labeled g, contained potassium iodide solution.[1] Five droppers, labeled to correspond with the bottles, were set up before each bottle. Two test tubes were filled with liquid to the same level ahead of time. One contained water and the other a mixture of diluted sulphuric acid and hydrogen peroxide. A rack of empty test tubes, paper and pencil and paper towels completed the set up.

The problem posed to the subject consisted of two stages. First, the interviewer pointed to the two prepared test tubes, told the subject that each contained one or more of the liquids in bottles one through four and then demonstrated the effect of g by adding several drops of g to each test tube. Since hydrogen peroxide oxidizes potassium iodide in an acid medium, the mixture 1 x 3 x g turned yellow while the mixture 2 x g (water and potassium iodide) remained colorless. After the subject had identified the effect of adding g, the interviewer completed the problem statement by

---

[1]The Inhelder reference (1958) specifies oxygenated water rather than hydrogen peroxide. The combination 3 x g (hydrogen peroxide solution and potassium iodide solution) may result in a "hint of a tint," but this is not easily confused with the demonstrated and expected color change.

asking the subject to reproduce the yellow effect by using liquids from the bottles ("any of them") and g.

A subject whose spontaneous behavior is limited to combining 1 x g, 2 x g, 3 x g and 4 x g or, perhaps, 1 x 2 x 3 x 4 x g, regardless of subsequent cued behavior, was classified as 2A. If the subject spontaneously (1) tried out other combinations (e.g., 1 x 2 x g, 3 x 4 x g) but unsystematically, or (2) gave evidence of using a systematic approach but was unable to use the results in formulating conclusions, he or she was classified as 2B. In contrast, the 3A subject spontaneously used a systematic approach and subsequently demonstrated the ability to make deductions on the basis of the evidence. According to Inhelder and Piaget, 3B subjects are distinguished from 3A primarily by the quality of the reasoning and the kind of system employed. Complete spontaneity in formulating a plan, systematically carrying out that plan and reasoning (deductively) toward necessary conclusions is required of the 3B subject.

The data and analysis which follow illustrate the typical "low" and "high" 2B and 3A subjects classified in the course of this study.

"Low" 2B subjects did spontaneously combine more than one liquid with g, but, whether successful in obtaining yellow or not, were extremely limited in awareness of all

the possibilities. Cues from the experimenter were singularly unsuccessful in extending the responses of these subjects.

Alyson (14 years, 1 month): [Combined 1 x 2 x g first, then tried 1 x 3 x g and got yellow.] "There it is." Yes, you're right. Perhaps there are other ways of getting yellow. "Maybe." [Tried 3 x g, 4 x 3 x g, 2 x 3 x g.] "No, that was the only way." Are you sure? Did you try everything possible? "Yes."

Alyson, like all the "low" 2B subjects in the sample, was unable to imagine any further possibilities. The possibilities which she did try after being prompted seemed to be based on the idea that 3 was instrumental in provoking the color since, with one exception, 3 appeared in each combination tested. The sole difference between the behavior of these "low" 2B subjects and the 2A subjects was that initial, spontaneous combining of more than one liquid with g.

The response of the "high" 2B subjects in our study, at first glance, appeared congruent to that of the better 3A subjects. The subjects typically proceeded systematically through all of the possible combinations and sometimes recorded part or all of the system on paper. The responses of Diane are an extreme example of this kind of behavior.

Diane (14 years, 7 months): [Recorded bottle numbers and results; i.e., yes or no as she worked. Tried all one

at-a-time combinations, then the two-at-a-time. Got yellow with 1 x 3 x g.] "May I continue?" Yes, certainly. [Recorded as before and combined remaining two-at-a-time with g and all three-at-a-time, except 1 x 2 x 4, with g. Stopped, smiled and indicated that she'd finished.] I notice you recorded 'Yes' for more than one situation. What ones were those? [Consulted list.] "1 x 3 x g and 1 x 2 x 3 x g." Can you explain that? "No, they just both work."

Further cues to see if Diane could deduce the role of 4 were equally unproductive. This example of behavior which is not only systematic, but painstakingly so, is a vivid illustration of the difference between the ability to write or list a set of combinations and the ability to use combinatorial methods. Moreover, like other "high" 2B subjects, Diane appeared to be aware that there might be more than one way to obtain yellow but exhibited no curiosity as to the cause of the results and was unable to hypothesize as to a probable cause despite all the evidence from her tests.

The "low" 3A subjects moved through the one-at-a-time and two-at-a-time combinations in a systematic manner until obtaining the yellow color with 1 x 3 x g. When asked if there was any other possibility, they immediately responded by continuing to recite and/or try several combinations, but not all possible combinations, in a systematic way. In

fact, they typically required strong prompts before giving convincing evidence of command of the combinatorial system.

Kevin (14 years, 8 months): [Tried 1 x 2 x g, 1 x 3 x g, got yellow.] Do you think there are other ways to get the color? "Yeah, I suppose I could try some other things." [Mixed 2 x 4 x g and 2 x 3 x g.] "I don't think 2 helps to get the color." Have you tried them all? "No." [Tried 1 x 4 x g and 3 x 4 x g.] Are there any other possibilities? "Oh, yeah, three together." [Tried all three-at-a-time combinations systematically.] "I should have written these down, I'm losing track." [Never did repeat or omit any combination, but did not try four-at-a-time.] What have you learned from these tests? "1 x 3 x g give the color. Two does nothing." Why are you so sure? "Look at 1 x 2 x 3 x g. It's the same as 1 x 3 x g." What does 4 do? "I think it wipes it out."

Although Kevin required repeated prompting to use additional combinations, he was convinced that the color was due to a combination, rather than to particular elements. Moreover he was, in fact, employing a systematic approach and his overt insecurity was a sign of his awareness of the potential complexity of the system--an awareness not possessed by most 2B subjects. Yet he was far less secure and confident than Diane, the "high" 2B subject, in demonstrating the combinatorial system. We found this kind of

imprecision in demonstrating full command of the combinatorial system among all "low" 3A subjects who, at the same time, were able to reason from the concrete evidence. All were able to use propositional logic with varied degrees of effectiveness and tried to reason from the data throughout the testing. Notice Kevin's early conjecture as to the role of 2. However, "low" 3A subjects did not use these early conjectures to predict subsequent results nor to structure the sequence of subsequent tests.

"High" 3A subjects were all characterized by clear control of the entire set of combinations. Ken began by trying all of the one-at-a-time combinations while keeping track with paper and pencil.

Ken (14 years, 9 months): "Could there be more than one?" There might be. [Tried 3 x 4 x g, 3 x 2 x g, 3 x 1 x g. Got yellow.] Might there be another way to do this? "Yeah, there probably could be." Would there be a lot? "No, just a few." [Named remaining two-at-a-time combinations.] Try a few and see. [Tested 2 x 1 x g. Stopped.] "Oh, well, there could also be three-at-a-time or even all four." [Named all these. Continued testing all other two-at-a-time combinations and 1 x 2 x 3 x g. (Had stopped recording combinations and results.)] Is that the same color? "Yes, probably 2 just dilutes this and gives it the same color."

Then how could you find out the role of 4? "I could try all four." Go ahead. [Mixes 1 x 2 x 3 x 4.] Before you add g, what do you think will happen? "Well, maybe it will just take it away." [Combines g with 1 x 2 x 3 x 4. Bleaches out.] Now what do you think? "Well maybe 4 and 2 in combination cancel it out." How could you find out? "Try 4 with 1 and 3 and g."

Ken's behavior rather quickly shifted from slow, careful, hesitant statements to confident generalizations. He seemed to be grappling with the nature of the problem at first and clearly continued thinking ahead while mixing liquids. After first seeming to possess only a limited set of combinations, he quickly and, without prompts, identified all the possible combinations. He was not only able to reason from the data but did so spontaneously in the case of the role of 2. This possession of the complete combinatorial system and the appearance of speed and facility in the use of the system (Ken stopped recording when the system got more complicated.) seemed to be a characteristic of "high" 3A subjects. "High" 3A subjects also differed in the kind of reasoning they used. Like Ken, they not only reasoned from the data, but structured their subsequent tests on the basis of deductions. Notice that Ken chose to mix 1 x 2 x 3 x 4 x g after reasoning about the

effect of 1 x 3 x g and 1 x 2 x 3 x g. These subjects were also able to make reasonable predictions going beyond the data they had obtained empirically whereas "low" 3A subjects were generally restricted to reasoning from the data.

The Shadows Task. This task was designed to elicit reasoning strategies related to the formal operational schema of proportionality. Unlike two other original Inhelder tasks which require proportional reasoning (equilibrium in the balance and hauling weight on an inclined plane), the projection of shadows task departs from a model of physical equilibrium and uses, instead, a model which is highly geometric in nature. The equipment used in this study (see Figure 1) included rings with diameters of 3 cm, 6 cm, 9 cm and 12 cm which could be hung from nails placed at 2 cm intervals on a hanger board 100 cm in length. This hanger board was labeled from 0 to 100 and 100 to 0. Unlike the other two tasks used in this study, the projection of shadows task was clearly subdivided (see Inhelder, 1958) into four hierarchically arranged subtasks. Success on all four subtasks was required to attain 3B classification; success on the first three, but not the fourth task, yielded a 3A classification and so on.

The functioning of the equipment was demonstrated to the subject with one ring in place. Then the subject was told

that he or she would be asked to make predictions about shadows while the light was turned off and give reasons for the predictions. Then the light would be turned on so the subjects could check on the correctness of their predictions. In subtask one, a 3 cm and a 9 cm ring were placed on a common nail and the subject was asked to predict the relative size of the shadows. "Will the shadows be equal or will one be larger than the other? If unequal, which ring will cast the larger shadow? Why?" The subjects who correctly related the size of the shadows to the size of the rings were classified as, at least, 2A.

In subtask two, the experimenter showed two 9 cm rings and placed each on a different nail (i.e., at unequal distances from the light). The questions asked previously were repeated. Subjects who predicted shadow size of ring correctly and explained either that the shadow is larger for the ring further from the screen or that the shadow is larger for the ring closer to the light were classified as, at least, 2B. In accord with the original Inhelder protocols, subjects were allowed a second trial on this subtask if they predicted incorrectly on their first attempt. On the second trial, the interviewer (with the light out) moved the rings to two new locations. The potential for learning from the previous trial is an interesting aspect of the shadows task.

Unsuccessful subjects sometimes not only corrected the prediction but also focused on the inverse proportion (i.e., shadow size increases as distance from the light decreases) which underlies this task.

In subtask three, subjects were asked to place a 3 cm and a 9 cm ring on the hanger board so that the shadows would coincide or "cover each other exactly." An explanation was required before the light was turned on and a second trial was permitted for unsuccessful subjects.

Finally, subjects successful on the first three tasks were given the fourth subtask. The interviewer placed a 6 cm ring at 36 (or 32) and subjects were asked to place a 9 cm and a 12 cm ring on other nails in such a way so that all three shadows would coincide when the light was turned on again. They were also asked to explain why these choices were made. Success on this subtask was evidence of 3B reasoning.

As noted earlier, very few subjects from the original pool of 69 showed evidence of formal reasoning on this task. Therefore, although it was possible to screen out "low" and "high" 2B responses, we were unable to make such a distinction within the formal sample on this task.

Since we are concerned in this paper only with 2B and 3A reasoning, the data and analysis which follows begins with subtask two (two 9 cm rings on different nails).

The "low" 2B subjects were limited to responses based on the concrete experience immediately before them. They required two trials to succeed on this subtask and when asked to place two unequal rings so that their shadows coincided (subtask three), they explained choices in terms of what just happened.

Lisa (15 years, 1 month): [Trial two with two 9 cm rings at 46 cm, 78 cm from light.] "That one (46 cm) is larger." Why? "Because I remembered that before the one closer to the light was larger." [Task three: places 3 cm ring at 72 cm and 9 cm ring at 12 cm.] Why did you choose those places? "I noticed from before that I needed to get the 3 cm shadow much larger than the 9 cm in order to cover one another. I noticed from before that the 9 cm shadow got larger when you put it closer to the screen. Whoops! May I change it?" Yes. [Places 9 cm ring at 72 cm and 3 cm ring at 12 cm.] Why did you do that? "Because I forgot. It really was larger when it was closer to the light." [Light on. Trial two. Trial and error adjustment of the 3 cm ring.]

The effect of the immediate feedback of subtask two is obvious from this protocol. On the one hand, the subject's initial memory lapse made one wonder how long she would retain the final relationship she stated. On the other hand,

there was clear evidence that Lisa was trying to apply the concrete results observed in the first two subtasks to a new situation. Her reasoning was typical of all other "low" 2B subjects in that it was limited to the immediately observed, concrete experience. Her explanation differed in one respect from those of other "low" 2B subjects. She did speak in terms of an inverse proportionality (larger shadow, closer to the light) while most others at this level spoke in terms of direct proportionality (larger shadow, further from the screen).

In contrast, "high" 2B subjects typically predicted correctly on trial one when given subtask two (equal rings at unequal distances) and may have even alluded to the dispersion of light, although the explanation still connected larger shadow with ring further from the screen. On the third subtask, these subjects always placed the rings in the correct order and explained their choices in terms of a metrical relationship. The response of David was a lucid example of a search for such a relationship.

David (15 years, 8 months): [Places the 3 cm ring at 40 cm (60 from screen) and the 9 cm ring at 80 cm (20 from screen).] Why did you choose those places? "Well, 3 is 1/3 of 9 and 20 is 1/3 of 60." [Light on. Subject expresses surprise at the result.] "That can't be right!" Do you

think you could improve the situation? [Light out. Moves the 3 cm ring to 20 cm (80 from the screen).] Why did you make that change? "I put them 60 apart because the rings are 6 apart."

David's concentration on distance from the screen, despite an earlier non-verbal allusion to a light cone and its role, led him to formulate a proportion which did not work. His overt dismay when the light was turned on indicated how sure he was of the predicted mathematical model. Yet he failed to consider other possibilities for applying the same model and, instead, tried out a different model which was additive in nature. Less capable 2B subjects typically used an additive model from the start and then tried to modify results by trial and error. However, "high" 2B subjects, like David, tried to extend the reasoning used for the earlier subtasks where there had appeared to be a ratio between shadow size and distance. They, too, spoke of the fact that the 9 cm ring was three times larger than the 3 cm ring but, like David, focused on the direct proportion they believed they had observed in action. Their initial search for a metrical relationship placed them close to the behavior of 3A subjects. However, they were unable to resolve the contradiction between predicted relationship and effect and began to search, in a trial and error fashion, for other numerical relationships, without regard for the

results of earlier subtasks.

Although the sample size classified as 3A or 3B was very small, our analysis of the protocols of the subjects classified as formal yielded some tentative conclusions as to what might be called "low" 3A behavior. The script which follows is included primarily to demonstrate the contrast between "high" 2B reasoning and this example of 3A reasoning.

Deron (14 years, 2 months): [Placed the 3 cm ring at 34 cm and the 9 cm ring at 66 cm.] "Well, 3 is 1/3 of 9 and so the 3 cm ring should be 1/3 of the way from the light to the screen. Thirty-four is close enough. Then the 9 cm ring should be the same distance from the screen. But something doesn't seem right." [Light on.] Would you like to try again? "Yes." [Light out. Placed the 3 cm ring at 20 cm and the 9 cm ring at 60 cm.] Why did you choose those places? "The 9 cm ring is three times the 3 cm so I just have to be sure the 9 cm ring is three times as far from the light as the 3 cm ring."

Unlike the "high" 2B subject, David, Deron was unwilling to part with the mathematical model which seemed consistent with earlier tasks. Deron also showed an understanding of the essentially geometric nature of the shadows tasks and seemed to be using a three-dimensional light cone in his reasoning. Notice, however, that his first trial

contained compound errors. He compared the rings, used that ratio to place the smaller ring, but then appeared to ignore ratio and use a concept of symmetry to position the 9 cm ring. Perhaps "low" 3A subjects will exhibit apparently contradictory reasoning while clearly searching for some kind of metrical relationship. However, it seems that their initial errors must always use an inverse relationship (larger shadow, closer to the light) and a multiplicative rule.

The Flexible Rods Task. The schema illustrated by a complete solution of the flexible rods problem is referred to as "all other things being equal" by Piaget. In this situation the subject is asked to identify all the variables which affect the flexibility of a rod. In addition, the subject must exhibit appropriate verification tests and verbalize the necessity for controlling all variables except the one being tested. Although other Inhelder tasks, such as the floating bodies problem, require the use of this schema, the flexible rods task is more difficult than these, primarily due to the existence of five distinct variables.

The equipment used in the study reported here consisted of five metal rods clamped to a backboard and a set of lead fishing sinkers (several identical small ones and several identical large ones) which could be hung on the ends of

the rods. The rods varied in terms of thickness (thin or thick), cross-sectional form (square or circular) and material (brass or steel). The length that each rod projected beyond the backboard could be varied by adjusting the clamp and pulling or pushing the rod.

Each subject was shown the equipment with rods projecting different distances from the backboard. Then one weight was placed on the end of one rod so that the rod bent far enough for the weight to barely touch the baseboard. The subject was shown the other weights and the way in which rod length could be changed. Then the problem was stated: "Identify all the things which make one rod bend more than another" or "make one rod bend a little or a lot." (The initial reaction of individual subjects to the verbalized situation sometimes led to a more sophisticated phrasing of the problem.) The subjects were then immediately reminded of the presence of paper and pencil and told to work with the equipment before verbalizing ideas. After some trial-and-error hands-on work, subjects were asked to identify variables and to show a test which would be convincing evidence for the stated effect of each identified variable. After such tests, subjects were asked why they had manipulated the equipment in a particular way, if any other choice of rods could have been used, and so on.

Subjects were classified as 2A if some of the relevant variables were systematically identified from the observations of the raw data. However, 2A students not only failed to understand the need for controls but also were unaware of the compensating effect of two relevant homogeneous variables. For example, 2A subjects not only might confound width with weight but also might compare a thin and long rod with a thick and short rod to test first for width ignoring completely the effect of length. Next, the same two rods might be compared to test for length with no consideration given to width. The change from 2A to 2B reasoning is subtle for the only additional reasoning is the appearance of an understanding of the compensating effect of two homogeneous variables. For example, 2B subjects may compare the same two rods referred to above but be aware that if the rods are of the same material then a wider rod may bend as much as a thinner rod provided the first rod was longer than the second. However, 2B subjects are unable to verify the effect of one factor by keeping all other known factors constant. The appearance of this schema, "all other things being equal," for even some of the identified variables was sufficient to classify a subject as 3A while the 3B classification was reserved for those subjects who identified all five variables and systematically demonstrated the use of controls in verification.

The data and analysis which follows describes distinct characteristics of "low" and "high" 2B and 3A subjects from this sample.

The "low" 2B subjects usually were able to identify a limited number of variables, typically the homogeneous ones, length, weight and/or thickness. In addition, although these subjects did explain some conclusions in terms of the compensating effects of two of these homogeneous variables, they also used trial and error extensively.

Susan (14 years, 3 months): [After some trials, identified length and thickness as factors which made a difference.] What could you do to show that length does make a difference in the way the rods bend? [Hung two equal weights on two circular brass rods, one thick and one thin. Moved the rods back and forth until both bent about the same.] "See, this bends more if I move it out further." They seem to bend about the same the way you have them. "Well, that one is thicker so you have to pull it out further because it's stronger." Is there any other way you could show that length or thickness makes a difference? "No, I don't think so." Are there any other things which might make a difference? "Maybe the shape." Could you test to see if it does? [Subject confounds dramatically.] Any other possible factors? "No, I've found all of them."

Like all "low" 2B subjects, Susan identified few variables but did begin to apply what Inhelder and Piaget (1958, p. 51) call double-entry tables. However, her ability to use double-entry tables was extremely limited. Obviously her failure to consider more variables limited the size of the double-entry table. So Susan appeared to be only one step removed from 2A subjects and was less successful than many of them in the number of variables identified.

"High" 2B subjects, on the other hand, typically identified four of the five variables and were very systematic in the use of a double-entry table. Frequently, these subjects appeared to control. However, subsequent questioning revealed that the subjects, in attempting to compare the effect of homogeneous variables, had accidentally selected two rods which would have been used if controls had been planned.

Mike (15 years, 2 months): [Worked with equipment systematically changing lengths and trying weights.] "Okay, weight and the kind of metal make a difference. The more weight you use, the shorter the rod can be." How could you show that the kind of metal makes a difference? [Chose two rods of same width, length and form but different metal and used same weights on each.] What does that tell you? "That one doesn't bend as much." Could you have shown that with

any other rods? "No." [There was another pair which could have been used.] Are there any other factors that make a difference? "Length and how thick they are." How could you test for either of those? [Used two rods of same metal, different thickness and same form. Placed same weights on rods pulled out at different lengths.] What does that show? "Well, the thicker one has to be longer to bend as much as the thinner one." [No other variables offered in response to questions.]

Mike appeared to be approaching the task in a very systematic way. However, his early response made it clear that he had observed the compensating effect of weight and length. Later, length and width were identified as having compensating effects. Like all "high" 2B subjects in the sample, he identified four of the five variables. Though he appeared to control when testing for material, his inability to elaborate the control or to identify form as a variable are clear indications of the absence of the "all other things being equal" schema.

"Low" 3A subjects all exhibited a common characteristic, instability in the use of the "all other things being equal" schema. After controlling and elaborating controls for at least one variable, they would regress to tests in which they confounded other variables.

Mark (15 years, 10 months): [Worked with the equipment. Identified length and width.] How could you test for either of those? [Tested appropriately for both length and width.] Why did you make sure those rods were pushed in like that? "I wanted to compare the effect of the thick and thin rods so I had to make sure everything else, including length, was the same....The steel ones don't bend as far as those others." Show that. [Tested appropriately for material.] Are those the only rods you could have used? "No, you could show it with these two." [Indicated other correct pair.] Other variables? "The shape might make a difference." How could you find out? [Confounded by controlling for width and length but not material or weight.] Have you found all the variables? "Oh, well, different weight will make a rod bend more or less." [Tested but confounded with material and thickness.]

The difference from the "high" 2B was clear. Subjects did use and elaborate controls in some cases. However, as more variables were considered, these subjects seemed to be unable to mentally control the system of all the remaining variables.

In contrast, "high" 3A subjects frequently identified four or five of the relevant variables after working freely with the equipment. In testing their conjectures, these students systematically controlled for all identified

variables or all but one. The chief difference, then, from "low" 3A subjects appeared to be the greater stability of the schema.

Christine (15 years, 8 months): [Worked with equipment.] "I think length, the thickness or thinness and the metal make a difference." [Tested correctly for length and defended test.] How could you test for material? "Move those two until they are the same length." [Pointed to two thick circular rods, one brass and one steel.] "Now put the 6 weights on each." [Interviewer follows directions.] "See, they bend different amounts and everything is the same except metal." Could you have shown this with other rods here? "Oh, yes." [Subject selected two thin circular rods of brass and steel and put identical weights on each.] Now, how could you test for thickness? [Subject initially confounded, then corrected self quickly and tested correctly.] For some reason, you've avoided this rod in your tests. [Pointed to square rod.] "Well, it's square. Maybe shape makes a difference." How could you find out? [Tested correctly for shape.] "The round one bends more. Less metal in that rod." Now, let's see, you've listed four factors that make a difference in bendability. Are there any others? "No, I can't think of any."

This "high" 3A subject made one slip, but immediately corrected herself and systematically controlled in all other

tests. Furthermore, like all "high" 3A subjects, Christine was able to elaborate on the controls used and sought reasons for the behavior of the rods. We found that weight was omitted in several cases by the sample of 3A subjects and surmised that it might seem so obvious as to not need mention. Yet, weight was one of the first variables listed by 3B subjects so we could not discount this omission. Christine had to be prompted to include cross-sectional form. This variable was omitted about as many times as it was identified by the sample tested although prompts were frequently needed to elicit it. As noted earlier, the one significant feature of all "high" 3A subjects (versus "low" 3A subjects) was the consistent, systematic use of controls and the extended use of verbal explanations most of the time. Their behavior differed from 3B subjects only in the failure to spontaneously elaborate what Inhelder and Piaget call "a complete proof."

CONCLUSIONS

For each of the three tasks used in this study, the investigators identified large differences in the reasoning patterns both within the late concrete stage and within the early formal stage. These contrasts are delineated in the previous section and portray vividly the vast strides cognitive development takes in progressing through a substage

and the slight, but significant, barriers "at the brink" of a new stage. The "high" 2B subjects, on the chemical combinations task for example, were superficially more systematic than the "low" 3A subjects but demonstrably less capable of hypothetico-deductive reasoning. Similar behaviors were noted for each of the other tasks.

Since each of the tasks was used to identify a different aspect of formal operations, some of the reasoning patterns exhibited were highly task-specific. However, there were some common reasoning patterns across tasks for each of the four subgroups studied.

Table 1 about here

In Table 1, the contrasts between "low" and "high" became evident. The typical descriptors used to distinguish concrete from formal seem to be realized in the behaviors under "low" 2B whereas "high" 2B subjects reasoned in ways which appeared, at first glance, to be formal patterns. A closer examination indicated that the latter <u>never</u> were able to function in a hypothetico-deductive way. Notice the slight shift which was representative of the "low" 3A subjects. Sometimes they <u>did</u> reason deductively, but then they regressed. In many respects, their reasoning seemed to be limited to <u>ex post facto</u>

reasoning whereas the "high" 3A subjects were clearly "planning ahead" as they progressed through the tasks.

Since this study did not consider the responses of "intermediate" subjects, a developmental pattern throughout either the 2B or 3A stage cannot be outlined here. The dramatic differences within substages do not, of themselves, negate the concept of gradual development. However, they do suggest that the classifications of 2B or 3A may mask significant intellectual differences. The reasoning patterns described by task and substage revealed large differences in scope and kind of ability within the late concrete and the early formal stages. The subjects classified as transitional in this study ranged all the way from limited perception of the tasks and inflexibility in reasoning to the ability to go beyond the data and the use of problem-solving strategies. Yet, an examination of the demographics of this population yields no evidence that the sample was unique. In fact, the sample could best be described as including "typical" youngsters teachers are expected to instruct in most secondary school situations.

The existence of the diverse reasoning patterns found in this study is of significance for practitioners and researchers. If Piagetian diagnostic tests are used by practitioners to guide instruction, then descriptive statements

limited to classification as concrete versus the formal stage are clearly inadequate. At the secondary school and collegiate levels where an increase in the percent of 2B and 3A subjects has been generally found, it is important that instructional leaders be made aware of the potential range of abilities and the discrepant nature of the reasoning. The most promising avenue for research appears to be curriculum analysis. The most extensive and systematic work being done in this area is that of Shayer and his colleagues (1978) in Great Britain on science curricula. In the United States, Karplus and colleagues (1977) have been active in considering the implications of Piagetian psychology for areas of secondary school science. Elkind (1980) cited a great need for formal curriculum analysis research based on Piagetian psychology in all areas of secondary school curricula. The results of this exploratory study are further evidence that such curriculum analyses are long overdue and must be finely tuned to reflect the diverse reasoning patterns of transitional students.

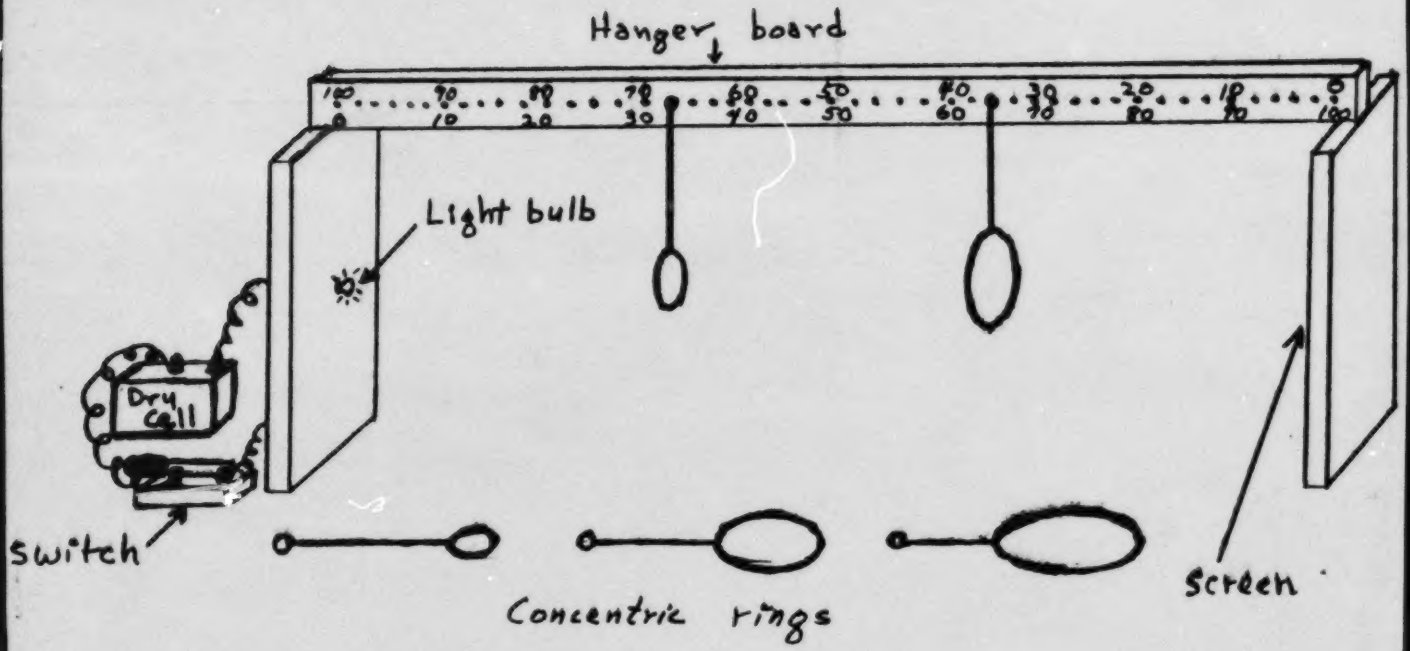

Figure 1: Apparatus for the Shadows Task

Table 1

Common Reasoning Patterns Across Three Inhelder Tasks by Subgroups

Low 2B	High 2B
Extremely limited ability to use concrete data immediately available	Aware of all or nearly all possibilities but unable to reason deductively from these
Explanations always in terms of concrete experiences immediately before them	Tries (unsuccessfully) to formulate metrical relationships
	Unable to resolve apparent contradictions
Initial signs of considering effect of two variables at one time	Considers effects of multiple variables but regresses as the number increases
Unsystematic approaches	Quite systematic
Cues/prompts add little to subject's repertoire	Cues/prompts extend subject's repertoire
Low 3A	**High 3A**
Instability in reasoning patterns (e.g., elaborates on controls in some instances, confounds in others)	Systematic approach and facility in deductive reasoning from the start, but requires prompts/cues to complete task
	OR
Little use of prior conjectures to structure further trials or to predict results	Progresses from slow, hesitant start to great facility in approach, reasoning and verbalization. (Seems to learn a lot from task.)

BIBLIOGRAPHY

Elkind, D. Investigating intelligence in early adolescence. In M. Johnson (Ed.), Toward adolescence: The middle school years. Chicago: The University of Chicago Press, 1980.

Farrell, M. A., et al. Piagetian diagnostic testing of senior high school mathematics and science students. Research paper presented at AERA Annual Meeting, Boston, Massachusetts, April, 1980.

Inhelder, B., and Piaget, J. The growth of logical thinking from childhood to adolescence. New York: Basic Books, Inc., 1958.

Juraschek, W. A. The performance of prospective teachers on certain Piagetian tasks. Unpublished doctoral dissertation, The University of Texas at Austin, 1974.

Karplus, R., et al. Science teaching and the development of reasoning. (Volumes for Biology, Chemistry, Earth Science, General Science and Physics). Berkeley, CA.: The University of California, 1977.

Lawson, A. E. The development and validation of a classroom test of formal reasoning. Journal of Research in Science Teaching, 1978, 15, 11-24.

Lawson, A. E. Letter to M. Farrell and W. Farmer, December 14, 1979.

McKinnon, J. W., and Renner, J. W. Are colleges concerned with intellectual development? American Journal of Physics, 1971, 39, 1047-1052.

Neimark, E. Intellectual development during adolescence. Child Development, 1975, 4, 541-594.

Neimark, E. Current status of formal operations research. Human Development, 1979, 22, 60-67.

Phillips, D. The Karplus islands puzzle, volume with metal cylinders, and other problems. Journal of Research in Science Teaching, 1977, 14(4), 361-365.

Pluta, R. F. The effect of laboratory and quasi-lecture modes of instruction on mathematical learning of formal and concrete operational college students. Unpublished doctoral dissertation, State University of New York at Albany, 1980.

Renner, J. W., and Stafford, D. G. Teaching science in the secondary school. New York: Harper and Row, 1972.

Shayer, M. The analysis of science curricula for Piagetian level of demand. Studies in Science Education, 1978, 5, 115-130.

Shayer, M. Has Piaget's construct of formal operational thinking any utility? British Journal of Educational Psychology, 1979, 49, 265-276.

Sheehan, D. The effectiveness of concrete and formal instructional procedures with concrete and formal operational students. Unpublished doctoral dissertation, State University of New York at Albany, 1970.